Scientific Reports, 2017, vol.7

Superstatistical model of bacterial DNA architecture

Bogachev M., Markelov O., Kayumov A., Bunde A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© The Author(s) 2017. Understanding the physical principles that govern the complex DNA structural organization as well as its mechanical and thermodynamical properties is essential for the advancement in both life sciences and genetic engineering. Recently we have discovered that the complex DNA organization is explicitly reflected in the arrangement of nucleotides depicted by the universal power law tailed internucleotide interval distribution that is valid for complete genomes of various prokaryotic and eukaryotic organisms. Here we suggest a superstatistical model that represents a long DNA molecule by a series of consecutive ~150 bp DNA segments with the alternation of the local nucleotide composition between segments exhibiting long-range correlations. We show that the superstatistical model and the corresponding DNA generation algorithm explicitly reproduce the laws governing the empirical nucleotide arrangement properties of the DNA sequences for various global GC contents and optimal living temperatures. Finally, we discuss the relevance of our model in terms of the DNA mechanical properties. As an outlook, we focus on finding the DNA sequences that encode a given protein while simultaneously reproducing the nucleotide arrangement laws observed from empirical genomes, that may be of interest in the optimization of genetic engineering of long DNA molecules.

http://dx.doi.org/10.1038/srep43034

References

[1] Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. Nature 421(6921), 423-427 (2003).

[2] Bryant, Z., Oberstrass, F. C. & Basu, A. Recent developments in single-molecule DNA mechanics. Curr. Opin. Str. Biol. 22(3), 304-312 (2012).

[3] Watson, J., Baker, T. A. & Bell, S. P. Molecular Biology of the Gene (7th Edition). (NY, Benjamin-Cummings Publishing Company, 2014)

[4] Guerra, C. F., Bickelhaupt, F. M., Snijders, J. G. & Baerends, E. J. Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. J Am. Chem. Soc. 122, 4117-4128 (2000).

[5] Arneodo, A. et al. Multi-scale coding of genomic information: From DNA sequence to genome structure and function. Physics Reports 498, 45-188 (2011).

[6] Li, W. & Kaneko, K. Long-Range Correlation and Partial 1/f Spectrum in a Noncoding DNA Sequence. Europhys. Lett. 17, 655-660 (1992).

[7] Peng, C.-K., Buldyrev, S. V., Havlin, S. et al. Long-range correlations in nucleotide sequences. Nature 356, 168-170 (1992).

[8] Peng, C.-K., Buldyrev, S. V., Havlin, S. et al. Mosaic organization of DNA nucleotides. Phys. Rev. E 49, 1685-1689 (1994).

[9] Buldyrev, S. V. et al. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. Phys. Rev. E 51, 5084-5091 (1995).

[10] Arneodo, A., Bacry, E., Graves, P. V. & Muzy, J. F. Characterizing long-range correlations in DNA Sequences from wavelet analysis. Phys. Rev. Lett. 74, 3293-3296 (1995).

[11] Sandberg, R. et al. Capturing whole-genome characteristics in short wequences using a naive-bayessian classifier. Genome Res. 11, 1404-1409 (2001).

[12] Hao, B. & Ji, Q. Prokaryote phylogeny without sequence alignment: From avoidance signature to composition distance. J. Bioinf. Comp. Biol. 2, 1-19 (2004).

[13] Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive bayessian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. J. Bioinf. Comp. Biol. 2, 1-19 (2004).

[14] Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R. Accurate taxonomy assignments from 16S RNA sequences produced by highly parallel pyrosequences. Nucl. Acids Res. 36, e120 (2008).

[15] Chor, B., Horn, D., Goldman, N., Levy, Y. & Massingham, T. Genomic DNA k-mer spectra: models and modalities. Genome Biol. 10, R108 (2009).

[16] Kuksa, P. & Pavlovic, V. Efficient alignment-free DNA barcode analysis. BMC Bioinformatics 10, 59 (2009).

[17] DeSantis, T. Z. et al. Simrank: Rapid and sensitive general-purpose k-mer search tool. BMC Ecology 11, 11 (2011).

[18] LaRosa, M., Fiannaca, A., Rizzo, R. & Urso, A. Alignment-free analysis of barcode sequences by means of compression-based methods. BMC Bioinformatics 14, S4 (2013).

[19] LaRosa, M., Fiannaca, A., Rizzo, R. & Urso, A. Probabilistic approach modeling for the analysis and classification of genomic sequences. BMC Bioinformatics 16, 52 (2015).

[20] Grosberg, A., Rabin, Y., Havlin, S. & Neer, A. Crumpled globule model of the three-dimensional structure of DNA. Europhys. Lett. 23, 373-378 (1993).

[21] Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293 (2009).

[22] McNally, J. G. & Mazza, D. Fractal geometry in the nucleus. EMBO J. 29, 2-3 (2010).

[23] Mirny, L. The fractal globule as a model of chromatin architecture in the cell. Chromosome Res. 19, 37-51 (2011).

[24] Schram, R. D., Barkema, G. T. & Schiessel, H. On the stability of fractal globules. J. Chem. Phys. 138, 224901 (2013).

[25] Tamm, M. V., Nazarov, L. I., Gavrilov, A. A. & Chertovich, A. V. Anomalous diffusion in fractal globules. Phys. Rev. Lett. 114, 178102 (2015).

[26] Bednar, J. et al. Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. J Mol. Biol. 254, 579-594 (1995).

[27] Vologodskaia, M. & Vologodskii, A. Contribution of the intrinsic curvature to measured DNA persistence length. J Mol. Biol. 317, 205-2013 (2002).

[28] Geggier, S., Kotlyar, A. & Vologodskii, A. Temperature dependence of DNA persistence length. Nucl. Acids Res. 39(4), 1419-1426 (2011).

[29] Bogachev, M. I., Kayumov, A. R. & Bunde, A. Universal internucleotide statistics in full genomes: A footprint of the DNA structure and packaging PLoS One 9, e0112534 (2014).

[30] Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. 52, 479-487 (1988).

[31] Renyi, A. On measures of entropy and information. Proc. 4th Berkeley Symp. Math. Stat. Prob. 1, 547-561 (1961).

[32] Grassberger, P. & Procaccia, I. Dimensions and entropies of strange attractors from a fluctuating dynamics approach. Physica D 13, 34-54 (1984).

[33] Nauenberg, M. Critique of q-entropy for thermal statistics. Phys. Rev. E 67, 036114 (2003).

[34] Presse, S. Nonadditive entropy maximization is inconsistent with Bayessian updating. Phys. Rev. E 90, 052149 (2014).

[35] Ludescher, J., Tsallis, C. & Bunde, A. Universal behaviour of interoccurrence times between losses in financial markets: An analytical description. EPL 95(6), 68002 (2011).

[36] Ludescher, J. & Bunde, A. Universal behavior of the interoccurrence times between losses in financial markets: Independence of the time resolution. Phys. Rev. E 90(6), 062809 (2014).

[37] Tsallis, C. Inter-occurrence times and universal laws in finance, earthquakes and genomes. Chaos, Solitions and Fractals 88, 254-266 (2016).

[38] Tamazian, A., Nguyen, V. D., Markelov, O. A. & Bogachev, M. I. Universal model for collective access patterns in the Internet traffic dynamics: A superstatistical approach. EPL 115, 10008 (2016).

[39] Beck, C. Dynamical foundations of nonextensive statistical mechanics. Phys. Rev. Lett. 87, 180601 (2001).

[40] Beck, C. & Cohen, E. G. D. Superstatistics. Physica A 322, 267-275 (2003).

[41] Naimark, O. B. Structural-scaling transitions and localized distortion modes in the DNA double helix. Phys. Mesomech. 10(1), 33-45 (2007).

[42] Touchette, H. & Beck, C. Asymptotics of superstatistics. Phys. Rev. E 71(1), 016131 (2005).

[43] Audit, B. et al. Long-Range Correlations in Genomic DNA: A Signature of the Nucleosomal Structure. Phys. Rev. Lett. 86, 2471-2474 (2001).

[44] Audit, B. et al. Long-range Correlations between DNA Bending Sites: Relation to the Structure and Dynamics of Nucleosomes. J Mol. Biol. 316, 903-920 (2002).

[45] Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S. & Bunde, A. Detecting long-range correlations with detrended fluctuation analysis. Physica A 295, 441-454 (2001).

[46] Goodsell, D. S. & Dickerson, R. E. Bending and curvature calculations in B-DNA. Nucl. Acids. Res. 22(24), 5497 (1994).

[47] Wiggins, P. A., et al. High flexibility of DNA on short length scales probed by atomic force microscopy. Nature Nanotechnol. 1(2), 137-141 (2006).

[48] Mazur A. K. & Maaloum, M. DNA flexibility on short length scales probed by atomic force microscopy. Phys. Rev. Lett. 112(6), 068104 (2014).

[49] Bresler, S. E. & Frenkel, Y. I. On the character of brownian motion of long organic chains and on the elastic properties of the rubber. J Exp. Theor. Phys. 9, 1094-1106 (1939).

[50] Kratky, O. & Porod, G. Röntgenuntersuchung gelöster Fadenmoleküle. Rec. Trav. Chim. Pays-Bas. 68, 1106-1123 (1949).

[51] Vologodskii, A. & Frank-Kamenetskii, M. D. Strong bending of the DNA double helix. Nucl. Acids Res. 41(14), 6785-6792 (2013).

[52] Mandelbrot, B. B. Gaussian Self-Affinity and Fractals. (NY, Springer, 2002)

[53] Hu, K., Ivanov, P. C., Chen, Z., Carpena, P. & Stanley, H. E. Effect of trends on detrended fluctuation analysis. Phys. Rev. E 64, 011114 (2001).

[54] Voss, R. F. Evolution of long-range fractal correlations and 1/f noise in DNA base sequences. Phys. Rev. Lett. 68, 3805 (1992).

[55] Bunde, A., Eichner, J. F., Kantelhardt, J. W. & Havlin, S. Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. Phys. Rev. Lett. 94, 048701 (2005).

[56] Altmann, E. G. & Kantz, H. Recurrence time analysis, long-term correlations, and extreme events. Phys. Rev. E 71, 056106 (2005).

[57] Bogachev, M. I., Eichner, J. F. & Bunde, A. Effect of nonlinear correlations on the statistics of return intervals in multifractal records. Phys. Rev. Lett. 99, 240601 (2007).

[58] Bogachev, M. I. & Bunde, A. On the predictability of extreme events in records with linear and nonlinear long-range memory: Efficiency and noise robustness. Physica A 390, 2240 (2009).

[59] Schehr, G. & Majumdar, S. N. Universal order statistics of random walks. Phys. Rev. Lett. 108, 040601 (2012).

[60] Majumdar, S. N., Mounaix, P. & Schehr, G. Exact statistics of the gap and time interval between the first two maxima of random walks and Lévy flights. Phys. Rev. Lett. 111, 070601 (2013).

[61] Bar, A., Majumdar, S. N., Schehr, G. & Mukamel, D. Exact extreme-value statistics at mixed-order transitions. Phys. Rev. E 93, 052130 (2016).

[62] Bogachev, M. I. & Bunde, A. Universality in the precipitation and river runoff. EPL 97, 48011 (2012).

[63] Bogachev, M. I., Kayumov, A. R., Markelov, O. A. & Bunde, A. Statistical prediction of protein structural, localization and functional properties by the analysis of its fragment mass distributions after proteolytic cleavage. Sci. Rep. 6, 22286 (2016).